

# Scale dependence in energy poverty measurement: theoretical implications

**Aner Martinez-Soto**

`aner.martinez@ufrontera.cl`

University of La Frontera <https://orcid.org/0000-0001-5087-9859>

**Katherine Castro-Coliqueo**

Universidad de La Frontera

**Daniel Sanhueza Catalán**

University of La Frontera

**Yarela Saldias-Lagos**

Gabriela Mistral University

**Alexander Zipf**

Heidelberg University

---

## Research Article

**Keywords:** Energy poverty, Multidimensional indices, Scale dependence, Indicator sensitivity, Multiscale analysis

**Posted Date:** October 13th, 2026

**DOI:** <https://doi.org/10.21203/rs.3.rs-11331863/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

**Additional Declarations:** The authors declare no competing interests.

---

# Abstract

Energy poverty is increasingly measured through multidimensional indices to identify vulnerable populations and inform energy and social policies. Yet whether the phenomenon captured by these metrics remains consistent across spatial scales is unresolved. Here, using 24 energy poverty indices across six territorial scales, we show that spatial aggregation does not simply alter estimated values but fundamentally changes what is measured. Aggregation selectively contracts dimensional representation, substantially reduces the discriminatory capacity of several indices, and shifts dominant determinants from household climatic, economic and efficiency conditions towards territorial and energy-system characteristics. Stress tests further reveal markedly different responses among conceptually similar indices under identical perturbations. These findings challenge the assumption that higher-level estimates simply represent aggregated household deprivation. The analysis was enabled by the Heidelberg Energy Poverty Dataset (HEP-Data), a harmonized multiscale benchmark of 261,833 households that also provides a reusable resource for testing existing and newly developed energy poverty metrics.

## Introduction

Energy poverty (EP) is recognized as a multidimensional phenomenon that intersects housing conditions, energy systems, and socioeconomic vulnerability<sup>1</sup>. Traditionally associated with the inability to access sufficient and affordable energy services<sup>2–4</sup>, contemporary approaches emphasize that EP reflects a broader set of structural deficits, including limited access to modern energy infrastructure, low thermal performance of housing, and exposure to extreme climatic conditions<sup>5,6</sup>. Che et al. (2021) also link EP to the Sustainable Development Goals, particularly SDG 7 on affordable and clean energy, SDG 3 on good health and well-being, and SDG 10 on reduced inequalities, highlighting the need for integrated approaches<sup>7</sup>. This conceptual expansion has driven the development of new measurement indices aimed at capturing the interaction between social, technical, and environmental determinants<sup>8,9</sup>.

The conceptual expansion increased the methodological complexity of the field without establishing a common framework for comparative evaluation. Differences in variable selection, weighting schemes, aggregation techniques, normalization procedures, and threshold definitions directly influence classification outcomes<sup>10,11</sup>. Empirical evidence shows that applying different normalization procedures or cutoff criteria to the same data can produce divergent results<sup>12–13</sup>. In practice, the selection of an index often depends more on data availability than on systematic comparative assessments<sup>14</sup>. Moreover, indices are applied at heterogeneous scales (household, municipality, region, or country) using non-harmonized empirical datasets<sup>15,16</sup>, and frequently rely on restricted or confidential microdata<sup>15,17</sup>, which limits their replicability and structural comparability.

Taken together, methodological heterogeneity, scalar variation, and restricted data access have limited the development of coherent comparative evaluations under controlled conditions<sup>18–21</sup>. Most

comparative exercises rely on contextual applications using different datasets and specific territorial scopes<sup>22–24</sup>. For example, Dobrowolski et al. (2025) analyzed multidimensional indices across European Union countries, but data heterogeneity limited structural comparability<sup>25</sup>. Zhu et al. (2025) highlighted the difficulties of applying composite indices in urban contexts in China due to inconsistent data availability regarding thermal performance and energy expenditure<sup>26</sup>. The absence of harmonized inputs makes it difficult to separate methodological effects from contextual variability<sup>27–29</sup> and constrains the development of comparable benchmarks across measurement frameworks<sup>30–32</sup>. This limitation is particularly critical because aggregation does not merely reduce spatial resolution, but also suppresses the detection of household-level vulnerability, under-consumption, and thermal deprivation patterns that remain visible only at finer scales of analysis. Scale dependence in spatial analysis is a well-established phenomenon, documented since the seminal work of Gehlke & Biehl (1934) and formalized as the Modifiable Areal Unit Problem (MAUP) by Openshaw (1984)<sup>33,34</sup>. Recent evidence confirms that composite vulnerability indices produce inconsistent spatial patterns when constructed at different scales, with vulnerability designations switching in nearly 22% of comparisons across data resolutions<sup>35</sup> yet no study has systematically quantified this instability across composite energy poverty indices under harmonized empirical conditions.

his study aims to develop and apply a multiscale validation framework to assess whether energy poverty measurement remains structurally consistent across spatial scales, focusing on changes in dimensional representation, discriminatory capacity, sensitivity, and dominant determinants. To enable this evaluation under controlled and comparable data conditions, we developed the Heidelberg Energy Poverty Dataset (HEP-Data), an open harmonized multiscale benchmark comprising 261,833 household-level observations, aggregated through a five-level territorial proxy structure (263,094 records in total across the six territorial scales). Using this benchmark, 24 indices were reproduced through a hierarchical protocol of variable substitution and traceability and applied across six territorial aggregation scales: household, neighborhood, parish, municipality, province, and country. By explicitly distinguishing effects associated with methodological design from those induced by scalar aggregation, the framework allows us to assess whether aggregation merely modifies energy poverty estimates or structurally reconfigures the information captured by existing metrics. Beyond its application in this study, HEP-Data provides an open and reusable benchmark for reproducing and comparing existing indices and for evaluating and developing new energy poverty metrics under standardized conditions.

## Results

**Scalar asymmetry and multidimensional information degradation in EP indices.** The findings of this study show that the global assessment of EP is structured by a scalar fragmentation that influences the consistency of diagnostics depending on the level of aggregation employed. The 24 indices analyzed in this study have been applied across multiple countries and territorial contexts in the literature. For this analysis, the six territorial scales were grouped into three levels: micro (Household), meso

(Neighborhood, Parish, Municipality, and Province), and macro (Country), where the unit of analysis corresponds, respectively, to the household, the intermediate territorial units, and the country (Fig. 1).

A total of 70.8% of the indices analyzed in this study operate under a single-scale design, predominantly concentrated at the Micro level (62.5%), which preserves distributive sensitivity in samples ranging from approximately  $10^2$  to  $10^6$  households (Fig. 1a). In contrast, Macro-level frameworks (16.7%) reduce sample sensitivity by operating on national units (5 to 146 countries), resulting in a loss of intra-country resolution. Only 12.5% of the instruments (FGT, FPI-1, and MEPI-2) implement bottom-up algorithms that allow transitions from microdata to aggregated outcomes; however, even in these cases, aggregation introduces statistical compression and potential smoothing of extremes.

Figure 1b shows that the pattern of information degradation is not only statistical but also multidimensional. At the Micro scale, the highest density of variables is concentrated (aggregated total = 141) across seven dimensions, particularly Infrastructure and Housing (39), Economic Capacity and Costs (24), Efficiency and Technology (24), and Social Vulnerability (22). This density enables the parameterization of complex deprivation configurations at the household level. However, when scaling toward the Macro level, the dimensional structure contracts substantially: Infrastructure and Housing and Economic Capacity are reduced to 5 variables each; Efficiency and Technology has no associated variables; Use and Consumption declines to 3; Perception and Governance to 1; and Climate and Environment gains only marginal presence, while Social Vulnerability concentrates 18 variables. This erosion of variables structurally simplifies the measurement system: by simultaneously reducing territorial granularity and the number of dimensions considered, the capacity to identify non-economic deprivations and complex combinations of vulnerability that emerge only at local scales is diminished. Variables related to dwelling quality, energy efficiency, and thermal habitability show greatest reduction across scales (Infrastructure/Housing: 39 to 5 variables; Efficiency/ Technology: 24 to 0 variables at macro level).

These findings establish that scale determines the statistical configuration of the measured phenomenon. When the unit of analysis, sample size, and dimensional coverage vary simultaneously, the internal structures of indicators and their ability to represent multicausal interactions are altered. Aggregated indices operate with lower empirical resolution and reduced dimensional diversity, limiting the detection of combined deprivations and modifying the relative weighting of physical, economic, and social components. Vertical comparability reflects structural differences across scales, with dimensional coverage ranging from 141 aggregated variables (micro) to 48 (macro).

**Spatial aggregation alters the discriminatory capacity of energy poverty indices.** At the Country level, values range between 0.00 (FPI (1), HER - NHERPGG, FPI (2), CEPI (2)) and 0.88 (EPVI) for the same case study in Fig. 2.a. Intermediate indices such as FPPRI (0.02), MMEP (0.30) and BFP (0.67) occupy disparate positions within that same range, evidencing that the country's classification as more or less energy poor depends entirely on the index chosen.

At the Parish scale ( $n = 10$ ), 40.0% of the indices present a median below 0.3, while for the remaining 20.0% the 90th percentile exceeds 0.7, evidencing a concentration of values toward the upper end of the distribution in that subset of indices. On the other hand, the internal variation of each index varies irregularly even when evaluating the same cases, as observed in the FPPRI and HEPI indices, which have coefficients of variation (CV) of 2.59 and 3.00, respectively, at the municipal scale, in contrast with the MEPI (2) index (152.75) and FGT (79.93), which present greater diversity of cases. However, when evaluating the same indices across different scales, this behavior becomes more stable, as in the FGT index, present in the 6 aggregation scales, which shows average values ranging between 0.17 and 0.24 ( $CV\% = 14.48\%$ ). The same occurs with the hEP (2) index, which has an average variation across all its scales of only 0.04 ( $CV\% = 3.06\%$ ), showing that indices can maintain their trends despite data aggregation.

Regarding the binary indices shown in Fig. 2.b, at the Household scale, the seven dichotomous indices range between 9.6% (EPI, MIS) and 97.1% (TPR), with 2M at 11.2%, hEP(1) at 11.0%, LIHC at 17.1%, and MEPI(1) at 27.3%; evidencing a certain similarity among 4 of them (2M, MIS, EPI, and hEP(1)), which require an income deficit as a classification criterion, explaining their lower and more homogeneous prevalence.

From the Neighborhood scale onward, four indices collapse toward 0.0%: LIHC and MEPI(1) retain marginal presence of poverty (1.5% and 2.6%), MEPI(1) persists at 1.3% in Parish, and all reach 0.0% from Municipality through Country. The TPR index is the exception, remaining at 100.0% from Neighborhood through Country, against 97.1% at Household. This is explained by modal aggregation: each territorial unit takes the most frequent classification of its households. Since the six minority indices (9.6%-27.3%) have a majority of non-poor households, that category dominates the mode at nearly every higher scale; TPR, majority since Household (97.1%), does the opposite. Thus, six of the seven indices are only informative at the Household scale, and TPR loses discriminatory power by converging toward a near-universal classification.

**Dominant determinants of energy poverty measurement reorganize across scales.** The results of the sensitivity analysis show that the hierarchy of critical variables in the measurement of EP is not stable across scales, but rather systematically reorganizes according to the level of aggregation used (Fig. 3).

At the Micro scale ( $n = 261,833$ ), climatic exposure concentrates the greatest influence on the variability of the indices. Heating Degree Days (HDD) maintain the highest sensitivity coefficient across the three perturbation scenarios, with values of 5.53, 5.46, and 4.21 for  $\pm 5\%$ ,  $\pm 10\%$ , and  $\pm 20\%$ , respectively. Gross and net household income present a mean coefficient of  $-1.02$  across all perturbations, showing greater sensitivity under the larger variation scenario ( $\pm 20\% \approx -1.04$ ). Together, these coefficients indicate that vulnerability to energy poverty at the micro scale is shaped by the interaction between climatic pressure (5.07) and lower payment capacity ( $-1.02$ ) in relation to monthly heating expenditures (0.62), which are influenced by higher energy costs associated with fuel prices (0.65) and heating system efficiency ( $-$

0.79). This indicates that household economic capacity is not always sufficient to sustain the energy costs associated with greater climatic demands.

On the other hand, at the Meso scale ( $n = 1,469$ ), the sensitivity hierarchy confirms the predominance of climatic and spatial variables, although with magnitudes that are more dependent on the size of the perturbation. Cooling Degree Hours (CDH), Heating Degree Hours (HDH), and summer nighttime temperature reach values of 5.00 ( $\pm 10\%$ ), as does habitable area per capita (5.00). However, their magnitudes increase under smaller variations ( $\pm 5\%$ ), reaching a value of 10.00, and decrease under variations of  $\pm 20\%$  (mean = 2.81). This reflects the discrete structure of the index formulation, where these variables are transformed into scores based on thresholds relative to the regional average, so that a smaller perturbation can generate a shift between categories without larger perturbations producing proportional increases. On the other hand, distributive effects associated with the gender gap are also observed: total monetary poverty maintains a positive response, with mean coefficients of 5.00 ( $\pm 5\%$ ), 2.50 ( $\pm 10\%$ ), and 3.75 ( $\pm 20\%$ ), while the sex-disaggregated indicators are mainly activated in the  $\pm 10\%$  and  $\pm 20\%$  scenarios: female poverty presents a positive response (1.88), and male poverty presents a negative response ( $-1.88$ ). This asymmetry responds to the weighting scheme of the NHERP<sub>GG</sub> index, in which the gap between female and male poverty directly modifies the quantification of EP.

Finally, at the Macro scale ( $n = 117$ ), the waiting time for electrical connection presents the largest coefficient in magnitude ( $-5.56$ ) and maintains the same value across the three perturbation scenarios. Since the GEVI index is constructed through standardization, aggregation by pillars, and PCA, the sign of the coefficient is interpreted as a direction of sensitivity within the statistical structure of the index, and not as a causal relationship. On the other hand, household energy expenditure reverses the sign observed at the micro scale ( $-1.52$ ). While at the micro level higher expenditure was interpreted as being associated with financial pressure, at the macro level a reduction in expenditure is associated with higher EP, indicating restrictions in the capacity to sustain a minimum level of energy consumption. This sign change suggests that the interpretation of household energy expenditure is scale-sensitive, although it may also reflect differences in the functional structure of the indices used at each scale. Variables associated with the composition of the energy matrix maintain the same sign and similar magnitudes across the three perturbation scenarios. Energy from fossil fuels presents a positive coefficient (1.31), indicating that greater fossil dependence increases energy vulnerability, while electricity generated from renewable sources shows a negative coefficient ( $-0.97$ ), reflecting a reduction in EP. This suggests that, at the macro scale, the national energy profile constitutes one of the main sensitivity factors of the indices.

**Metric architecture governs responsiveness under common stress conditions.** The stress analysis reveals clear quantitative differences in the ability of indices to discriminate between scenarios (Fig. 4). At the micro scale, 2M and TPR cover the full spectrum (0–1), but exhibit low mean values (0.046 and 0.004, respectively), indicating that although they capture extreme cases, most observations are concentrated near zero. hEP(2) reaches a range of 0–0.646 with a mean of 0.036, also maintaining relatively low dispersion. EPR shows substantial variation (0.019–0.568), whereas FPPRI remains

constrained between 0.014 and 0.055. At micro scale, 2M and TPR span full range 0–1 (means 0.046, 0.004); hEP(2) ranges 0–0.646 (mean 0.036); EPR shows 0.019–0.568 amplitude; FPPRI confined to 0.014–0.055. FPPRI incorporates 10 variables; EPR uses 8.

At the meso scale, EPVI spans the full range (0–1), demonstrating sensitivity to both increases and decreases in energy demand. Results obtained with FPI (1) fluctuate between 0.500 and 0.699 (mean 0.595), whereas FPI (2) remains in the lower quartile ( $\leq 0.25$ ; mean 0.127) even under extreme scenarios. HER–NHERP, GG also covers 0–1, but maintains a mean of 0 under the reference scenario, and BFP remains constant around 0.336 with no variation across scenarios. These results indicate that theoretical coverage of the spectrum does not always translate into effective shifts in the distribution.

FGT (amplitude 0–0.974, mean 0.248) and MEPI(2) (constant 0.333 across scenarios) are both expenditure-based. CEPI(1) reaches maximum 0.375; CEPI(2) minimum 0.368, with approximately equal amplitude. Discriminatory capacity varies across indices. Micro: 2M and TPR span 0–1; hEP(2) 0–0.646. Meso: EPVI spans 0–1; FPI(1) 0.500–0.699; FPI(2)  $\leq 0.25$ ; HER–NHERP spans 0–1 (mean 0 reference); BFP invariant 0.336. Macro: FGT 0–0.974; MEPI(2) invariant 0.333; CEPI(1) and CEPI(2) amplitudes  $\sim 0.343$  (offsets 0.375 vs. 0.368).

## Discussion

**Theoretical evolution and methodological heterogeneity in energy poverty metrics.** The evolution of energy poverty (EP) research reflects a gradual transition from approaches centered on economic deprivation and access toward multidimensional and energy justice frameworks. As noted by Guevara et al. (2025), the field has historically developed through five partially independent streams of thought: economic deprivation, access and affordability, capability-based approaches, vulnerability frameworks, and energy justice perspectives, each shaping distinct methodological trajectories and indicator formulations<sup>36</sup>. Kajoskoski et al. (2025) show that European research on energy poverty has evolved from economic quantifications of fuel poverty toward structural and intersectional perspectives linking energy vulnerability with broader social inequalities<sup>37</sup>.

This theoretical plurality explains the methodological heterogeneity observed among indices, where differences in threshold definitions, dimensional weighting, and aggregation procedures reflect divergent conceptual foundations rather than empirical inconsistencies, even when analyzing the same case study<sup>38</sup>. At the same time, the convergence toward an energy justice perspective highlights that the methodological fragmentation of current indices reflects a deeper epistemological diversity within the field, in which each metric operationalizes a specific theoretical stance regarding equity, need, and capabilities.

In this context, the framework developed in this study provides a controlled testing environment: by standardizing 23 diverse indices, it identifies the structural breaking points where energy poverty metrics

lose their consistency across scales, ensuring that policy decisions are based on stable evidence rather than statistical artifacts.

**Scalar dependence as a structural property of energy poverty measurement.** The scale-dependent restructuring documented here is consistent with a long-standing tradition in spatial science. Gehlke & Biehl (1934) first demonstrated that correlation coefficients increase systematically under spatial aggregation<sup>33</sup>; Robinson (1950) showed that aggregate-level correlations need not correspond to individual-level relationships<sup>39</sup>, what became known as the ecological fallacy; and Fotheringham & Wong (1991) extended these findings to multivariate analysis, showing that parameter estimates from regression models are highly sensitive to the spatial unit of analysis<sup>40</sup>. More recently, studies of composite vulnerability indices have confirmed that scale and indicator selection shift spatial patterns and reshape the relative importance of individual indicators<sup>35</sup>, and that poverty determinants exhibit higher spatial dependence at aggregated administrative levels<sup>41</sup>. The contribution of the present study is not the identification of scale dependence as a phenomenon, but its first systematic quantification across 23 composite energy poverty indices under harmonized empirical conditions, and the development of different response amplitudes under the same stress conditions as a formal validation dimension for index robustness across scales. Scalar dependence as a structural property of energy poverty measurement. Energy poverty (EP) indices systematically decrease their sensitivity as aggregation increases, confirming a scale-dependent attenuation of methodological responsiveness. This pattern is consistent with the heterogeneous  $\beta$ -convergence observed by Huang et al. (2022), where convergence rates in energy poverty vary across quantiles and structural contexts<sup>42</sup>. The phenomenon itself restructures at each scale. At household level ( $n = 261,833$ ), thermal exposure dominates (HDD coefficient 5.46), constrained by income ( $-1.01$ ) and efficiency ( $-0.79$ ), a direct physical interaction where climate intensifies demand, income limits spending, and technology modulates consumption. At municipal level ( $n = 1,469$ ), the same three factors persist (HDD 5.00, habitable area 5.00) but gain gender-differentiated effects ( $\pm 2.5$ ), revealing distributive disparities invisible at individual level. At national level ( $n = 117$ ), thermal variables recede and income dynamics invert ( $-1.52$  vs.  $+1.01$  at micro), replacing household mechanisms with structural constraints on minimum consumption. This scalar reorganization means that energy poverty is fundamentally differently constituted at each scale. These differences have critical implications for measurement and index selection. Indices with greater elasticity, those based on thermal comfort thresholds, maintain discriminatory capacity at household scales but lose sensitivity under aggregation, becoming non-comparable across levels. In contrast, metrics with constrained aggregation schemes (MEPI-TOPSIS CV = 13%; FPI(1) CV = 20%) show greater stability but lower local differentiation. From a measurement perspective, this implies that index selection should depend explicitly on the scale of intervention: elastic indices for detecting household-level thermal deprivation and retrofit needs, stable metrics for national benchmarking and long-term policy monitoring. Notably, low measured expenditure at household level may reflect constrained consumption rather than low vulnerability, a distinction that disappears at macro scales and is particularly relevant for identifying hidden energy poverty, where households with inadequate indoor thermal conditions remain invisible precisely because their energy expenditure is suppressed

**Implications for hidden energy poverty and residential thermal deprivation.** The disappearance of efficiency and housing quality variables at macro scales has direct implications for identifying hidden energy poverty. Households with low measured expenditure may not reflect low vulnerability but constrained consumption driven by limited disposable income. At household level, thermal efficiency (coefficient – 0.79) and climatic exposure (HDD 5.46) are critical predictors, yet neither features in national aggregates. Variables showing greatest reduction across scales include Infrastructure/Housing (39 to 5), Efficiency/Technology (24 to 0 at macro), and heating efficiency measures. This measurement gap implies that macro-level indices cannot identify the combination of low consumption plus poor indoor conditions that characterizes hidden deprivation. Comprehensive measurement thus requires integrating modelled thermal demand, actual consumption, indoor environmental conditions, and residual income dimensions systematically suppressed through aggregation

**Strategic implications for public policy, planning, and sustainable development.** From a public policy perspective, the measurement patterns revealed under the benchmark conditions provide evidence on how metric selection and analytical scale can influence the identification of energy vulnerability and, consequently, the information available for policy design. From this perspective, the findings suggest that macro-level indicators are appropriate for national benchmarking and long-term monitoring, meso-level indicators are more suitable for territorial prioritization and municipal targeting, while micro-level indicators are indispensable for detecting household-level thermal deprivation, constrained consumption, and retrofit needs. Consistent with the multidimensional link between EP and social well-being demonstrated by Liu et al. (2022), it has been observed that reducing EP contributes to lowering health risks, educational disruptions, and income deprivation, thereby reinforcing broader social benefits aligned with Sustainable Development Goals 1 and 7<sup>43</sup>. Likewise, the moderating role of EP in the outcomes of environmental regulations, evidenced by Tang et al. (2024), suggests that the welfare effects of sustainability policies depend on the preexisting level of energy deprivation: stricter regulations may exacerbate inequalities in contexts where energy access remains limited, whereas more flexible regulatory frameworks tend to improve welfare when accompanied by targeted energy reforms. These findings highlight that coherence between environmental and social objectives depends on the integration of energy poverty metrics into regulatory design<sup>44</sup>. Complementary evidence from He et al. (2025) reinforces this perspective by showing that the Provincial Clean Energy Demonstration Policy in China significantly reduced EP through processes of structural adjustment and improvements in energy efficiency, while strengthening synergies between pollution reduction and carbon mitigation<sup>45</sup>. This empirical pattern supports the argument that the promotion of clean energy and institutional design can simultaneously address energy vulnerability and environmental sustainability when policies are spatially differentiated and based on specific mechanisms.

**Stress Test Findings and Index Robustness.** Current validation frameworks emphasize theoretical coverage and variable count as robustness criteria<sup>46</sup>, yet variable count alone does not determine discriminatory capacity. Although FPPRI (10 variables) and EPR (8 variables) use comparable inputs, their elasticities differ dramatically (FPPRI  $\Delta = 0.041$  vs. EPR  $\Delta = 0.549$ ). Similarly, FGT and MEPI(2) both

use expenditure-to-income ratios but produce opposite responses: FGT spans full range (amplitude 0–0.974, mean 0.248) while MEPI(2) remains invariant (0.333), and CEPI(1) and CEPI(2) classify observations at opposite positions (0.375 vs. 0.368) despite similar amplitudes. These patterns reveal that methodological choices, threshold definition, normalization, internal structure, not data volume determine discriminatory capacity. While recent energy poverty studies conduct sensitivity analyses on weighting<sup>47</sup>, our results demonstrate that these methodological choices, independent of variable count, fundamentally reshape index behavior under perturbation. Similar to climate health vulnerability findings where design choices (inductive vs. deductive aggregation) significantly affect identification of vulnerable populations<sup>48</sup>, we establish different response amplitudes under the same stress conditions, as a critical but underutilized validation property. The divergence in European energy poverty frameworks where similar indicators produce disparate rankings<sup>49</sup> underscores the need for standardized robustness testing. We propose that future protocols explicitly quantify elasticity through systematic stress testing, providing a quantifiable basis for selecting indices aligned with analytical resolution required at specific scales and policy contexts.

**Limitations and Opportunities for Future Research.** This study's findings point to several areas where improvements in data availability and further research could advance the field. First, HEP-Data was designed as a harmonized benchmark for controlled methodological testing rather than as a population-representative survey. Its construction integrates observed data, proxy variables, and Approximate Synthetic Values (ASVs) to enable heterogeneous energy poverty indices to operate within a common multiscale data environment. Consequently, the results should be interpreted as evidence of metric behaviour, scalar dependence, and methodological responsiveness rather than as estimates of energy poverty prevalence in a specific population. This distinction is also central to the benchmark's intended use: HEP-Data provides a reproducible environment in which existing indices can be compared and newly developed metrics can be tested under standardized conditions. Future versions could expand the benchmark across alternative climatic, socioeconomic, and energy-system configurations to assess the generalizability of metric behaviour. Second, while our analysis examined energy poverty in Chile and its different climate zones, the patterns we identified reflect this specific geographic and socioeconomic context. Future research in other countries, with different climates, energy systems, and household economics, would reveal whether these same measurement challenges occur globally or are unique to certain regions. This geographic expansion represents a clear opportunity to strengthen our understanding of how energy poverty measurement works differently around the world. Third, our comparison at the national level includes 117 countries, but this number is limited by which countries systematically report energy data to international databases. As more countries develop and publish detailed energy statistics, researchers will have the chance to test whether our findings hold true across a more complete and representative sample of nations. These gaps in existing data highlight the importance of improving global energy information systems to support better policy decisions worldwide.

## Conclusions

This study demonstrates that the measurement of energy poverty is not structurally invariant with respect to the scale of aggregation. By comparing 24 indices applied to a harmonized empirical database of 261,833 households aggregated across six territorial scales, the results show that territorial aggregation simultaneously reconfigures the dimensional structure of the indices, the statistical coherence between metrics, and the hierarchy of determinants associated with the phenomenon. Taken together, these findings indicate that scale is not merely a descriptive property of the data but a structural condition that redefines the informational architecture through which energy poverty is observed and quantified.

The differences observed across scales also reveal that the discriminatory capacity of indices depends on their response amplitude under the same stress conditions. While some indicators retain sensitivity at local scales, others exhibit greater stability under aggregation, suggesting that different methodological approaches capture distinct analytical configurations of the phenomenon. Consequently, the selection of metrics should not be considered neutral with respect to the level of analysis, but rather explicitly aligned with the scale of intervention and the objectives of public policy.

More broadly, these results suggest that scalar dependence constitutes an inherent property of the measurement of complex socio-energy phenomena. The comparative framework proposed here provides a methodological basis for evaluating the coherence, sensitivity, and robustness of multidimensional indices under controlled conditions. Complementarily, HEP-Data provides an open and reusable multiscale benchmark for reproducing and comparing established energy poverty indices and for testing newly developed metrics under standardized conditions. Together, these contributions strengthen the methodological basis for selecting metrics according to analytical scale and policy purpose.

## Methodology

**Selection of energy poverty indices.** The selection of EP indices was conducted through a structured literature review covering the period 2010–2024. The search was performed in the Scopus and Web of Science databases using a multi-term query applied to titles, abstracts, and keywords: ("energy poverty" OR "fuel poverty" OR "energy vulnerability" OR "energy deprivation" OR "energy insecurity") AND (index OR indicator). The retrieved records were first cleaned by removing duplicates and subsequently filtered following the SALSA framework<sup>50</sup>, which provides a structured and replicable process for identifying, evaluating, and synthesizing scientific literature without the clinical constraints of health-oriented review protocols. Only studies meeting three criteria were included: (i) explicit methodological formulations enabling replication, (ii) quantifiable input variables derived from surveys or standardized databases, and (iii) prior validation in peer-reviewed scientific literature. Conceptual, normative, or non-replicable works were excluded.

The process resulted in 24 energy poverty indices derived from 23 publications, encompassing economic, multidimensional, and vulnerability-based approaches. These indices were developed across

different geographical contexts and analytical scales, ranging from household surveys to national-level indicators. These indices differ in their original entry and output scales, from household to neighborhood, parish, municipality, province, and country, as detailed for each index in Supplementary Table S1, together with their original sources. For the spatial aggregation analysis, EVCI and GEVI were excluded. EVCI was excluded because its calculation relies on subjective assessment criteria, preventing a standardized and reproducible implementation across the synthetic cases and aggregation levels. GEVI was excluded because its estimation requires cross-sectional variability among country-level units. Since the synthetic dataset represents a single country, aggregation to the national scale produces only one observational unit, resulting in zero between-unit variance and therefore precluding the comparative normalization and estimation required by the index.

**Construction of the Heidelberg Energy Poverty Dataset (HEP-Data).** The Heidelberg Energy Poverty Dataset (HEP-Data) was developed as a harmonized multiscale benchmark to provide a common data environment for the reproducible implementation, comparison, and testing of heterogeneous energy poverty metrics. Its purpose is methodological rather than population-representative: HEP-Data enables indices originally developed using different variables, datasets, units, and spatial scales to be implemented under standardized conditions. The construction process comprised two sequential stages: (i) collection and/or hierarchical substitution of data, with harmonization of units, scales, and reference periods, and (ii) systematic integration of variables into a unified multiscale structure.

For each selected index, the variables and input data defined in its methodological formulation were extracted, where variables refer to the conceptual components of the index (e.g., net income, energy consumption) and data refer to their numerical, categorical, dichotomous, or nominal values. This involved searching for the original data and variables reported in each index's corresponding source publication, following the definitions established in its methodological formulation. When original data were unavailable, proxies or Approximate Synthetic Values (ASVs) were incorporated following the hierarchical decision protocol shown in Fig. 5. ASVs were used to complete the standardized input architecture required by the benchmark when no directly compatible empirical observation or defensible proxy was available. Their purpose was to preserve the operational requirements of each index under controlled testing conditions rather than to reproduce the prevalence of energy poverty in a specific population.

Once the source data had been obtained, the process started from a household-level database (261,833 households) to which a sequence of corrections was applied to ensure the internal structural and temporal coherence of its variables, including income, household composition, tenure, housing expenditure, and the cascading derivation of energy expenditure and consumption by source. The consistency of the resulting database was validated through an audit of 18 rules, yielding a fully harmonized empirical database.

As a first step, prior to the construction and aggregation of variables described below, and to evaluate each territorial scale present in the indices, a nested hierarchical territorial clustering process organized

by thematic domains was carried out on the household-level database (MiniBatchKMeans), constructing a five-level territorial proxy structure (climate region, province, municipality, parish, and neighborhood), using the household sample's own climate region as the primary anchor. The process included a prior assessment of variability and redundancy (Pearson, Spearman, Cramér's V, Eta; threshold 0.60<sup>51</sup>). The final variables per domain corresponded to climatic and territorial characteristics (provincial level), housing and habitability (municipal level), energy and income (parish level), and household assets and social vulnerability (neighborhood level). Granularity was controlled through a maximum number of subdivisions per level and a minimum parent-unit size, via a sensitivity test at the neighborhood level (30 dwellings, 5% relative minimum), with 2 residual exceptions (0.02% of households). The clustering procedure yielded 38 provinces, 110 municipalities, 318 parishes, and 794 neighborhoods.

As a second step, all variables required for calculating the indices at the territorial scales above the household (Neighborhood, Parish, Municipality, Province, and Country) were constructed, starting from the household-level database. For indices whose original entry scale was the household, variables were aggregated directly from the microdata to each of the five higher scales (never chaining from an already-aggregated scale, to avoid bias from unequal group sizes), according to variable type: continuous and discrete variables via the mean; ordinal variables via the median; nominal variables via the mode; binary/dichotomous variables via a dual calculation of proportion and mode; and constant variables left unchanged. For indices with variables native to a scale above the household, these were incorporated at their native scale and scaled up to higher scales in two ways: on one hand, some variables were built as direct proxies from the microdata, using household-level anchor columns and their corresponding aggregation method (mean, ratio of sums, or constant); on the other, variables without a direct proxy were scaled against the real distribution of each index's native file, through percentile ranking on a theoretical household-level anchor column, subgroup-filtered ranking, or a composite score, depending on the nature of each variable. The entire process was audited variable by variable and empirically validated against the exported files, with no missing variables, no null values, and consistency across scales.

Of the 151 base variables that underpin the construction of the traceability matrix (49 household-level anchors, 42 native to the higher territorial scales, and 60 corresponding to a subjective index), 29.3% (44) correspond to real data, 16.0% (24) to proxy values, and 54.7% (82) to Approximate Synthetic Values (ASV). The databases constructed for this study are available at the following LINK: the first contains the variables of the indices native to the household scale, aggregated toward their respective higher territorial scales; the second contains the variables of the indices native to scales above the household (Neighborhood, Parish, Municipality, Province, and Country), with their corresponding scaling.

## Multiscale metric testing framework

*Application of Indices under Controlled Conditions.*

A common replication and simulation protocol was implemented using individual Excel calculators for each index and automated routines, in order to reproduce the original calculation procedures described in each source publication. The process began with the extraction of the calculation methodologies directly from the original references, taking into account the definitions and specificities of each variable. Based on these formulations, the procedures were faithfully replicated without structural modifications, so that each index preserved its internal logic and conceptual meaning in its corresponding calculation file. Each index was implemented independently, programmed to receive the input variables for the full set of cases at each territorial scale. The calculation was executed using Excel calculators and scripts assisted by artificial intelligence tools, under the authors' direct supervision and validation at each stage, which enabled the automation of results for each index under homogeneous conditions.

During the simulation, two types of results were distinguished: indices with binary results and indices with continuous values. In both cases, the original result type defined in the index's methodological formulation was preserved, in order to remain faithful to it without introducing structural modifications. However, the results of 4 continuous indices fell outside the standard 0 to 1 range; these cases were rescaled using the Min-Max normalization technique, defined in Eq. 1, in order to bring all indices onto the same comparable scale.

Eq. 1

$$Y_{norm} = \frac{Y - Y_{min}}{Y_{max} - Y_{min}}$$

To compare the behavior of the applied indices, an integrated results analysis was developed based on the distribution of each index's normalized values (0–1 range) across the six territorial scales, contrasting their internal and cross-scale variability through the coefficient of variation (CV%) and violin plots. For binary indices, the percentage of cases classified as energy poor was calculated at each scale. This procedure was programmed using automated routines for statistical comparison among indices.

## Sensitivity Analysis and Stress Testing

To evaluate the sensitivity of the indices to variations in their input variables, a univariate sensitivity analysis was conducted. The relative sensitivity was estimated by comparing the proportional change in the value of each index with the proportional change in the corresponding input variable, using finite differences in both terms. In this formulation, the numerator represents the relative variation of the index, while the denominator represents the relative variation of the perturbed input.

In this analysis, perturbations of  $\pm 5\%$ ,  $\pm 10\%$ , and  $\pm 20\%$  were applied to the continuous variables of indices with continuous outputs. For this analysis, the six territorial scales were grouped into three analytical levels: micro (Household), meso (Neighborhood, Parish, Municipality, and Province), and macro (Country). Each perturbation was applied independently, modifying one variable at a time while keeping the remaining variables constant. Sensitivity coefficients were calculated at the variable level

within each of these three levels, using averages to obtain one representative value per variable and scale. No upper bound, saturation rule, or capping procedure was imposed on the normalized sensitivity coefficients. This procedure made it possible to identify the most influential variables in the results of each index, facilitating the interpretation of the factors that affect their behavior<sup>52</sup>.

Finally, the Normalized Sensitivity Coefficient (NSC) was calculated to compare sensitivities among variables with different units or scales. Eq. 2 presents the expression used, where  $y$  is the value of the index,  $x_i$  is the evaluated input variable,  $\frac{\partial Y}{\partial x_i}$  is the partial derivative of the index with respect to that variable, and the entire term expresses the relative change in  $y$  given an infinitesimal relative change in  $x_i$ :

Eq. 2

$$S_i = \frac{\partial y}{\partial x_i} \bullet \frac{x_i}{y} = \frac{y(x_i + \Delta x_i) - y(x_i - \Delta x_i)}{2\Delta x_i} \bullet \frac{x_i}{y}$$

This metric is conceptually equivalent to relative sensitivity but expressed in terms of partial derivatives, which makes it particularly useful for comparative analysis among heterogeneous variables.

Subsequently, a stress test was conducted with the objective of evaluating the coherence and robustness of the indices under extreme conditions. All variables that presented values > 0 in their sensitivity coefficients were selected and, in each index, all of them were simultaneously modified, generating three scenarios: 1) Vulnerability to EP, characterized by high energy expenditures, low economic income, and inefficient housing; 2) No vulnerability to EP, associated with high income, low fuel prices, and low monetary poverty rates; and 3) Average case, which uses the mean value of all modified variables, acting as a reference line. The construction of these scenarios was based on the direction of the effect of each variable on EP. For the average case, the empirical mean of all input variables per index was used. For the extreme scenarios, it was previously identified whether each variable exerted a positive or negative effect on the level of energy poverty. In the vulnerability scenario, the variables were assigned the maximum observed values when their increase intensified EP (for example, energy expenditure), and the minimum values when their reduction aggravated vulnerability (for example, income). Similarly, in the non-vulnerable scenario, the maximum values were selected only for the variables whose increase reduced EP (for example, income), while the variables with an adverse effect adopted their minimum values. The scenarios were applied to the indices with continuous results, maintaining the original synthetic sample. This made it possible to record the variation induced in each index and to evaluate its internal conceptual coherence, also identifying which indices showed greater or lesser sensitivity to extreme conditions.

## Declarations

## Data availability

The datasets generated and analysed during the current study are available in the heiDATA repository (Heidelberg University) at DOI: 10.11588/DATA/FAWSVO

## Competing interests

The authors declare no competing interests.

## Author contributions

A.M.S.: Conceptualization, Methodology, Supervision, Project administration, Writing – original draft.

## Acknowledgements

A.M.S. acknowledges support from the Alexander von Humboldt Foundation through a Humboldt Research Fellowship, during which part of this research was developed.

## References

1. Lu S, Ren J (2023) A comprehensive review on energy poverty: definition, measurement, socioeconomic impact and its alleviation for carbon neutrality. *Environ Dev Sustain*. <https://doi.org/10.1007/s10668-023-04143-7>
2. Tundys B, Bretyn A, Urbaniak M (2021) Energy poverty and sustainable economic development: An exploration of correlations and interdependencies in european countries. *Energies* (Basel). 14
3. Boardman B (2013) *Fixing Fuel Poverty. Fixing Fuel Poverty: Challenges and Solutions*. Routledge. 10.4324/9781849774482
4. Galvin R (2022) Radically reducing UK energy poverty by the 10% and LIHC indicator through progressive fiscal policy: What would it cost, who would pay, and what are the consequences for CO2 emissions? *Sci Talks* 4
5. Herrero ST (2017) Energy poverty indicators: A critical review of methods. *Indoor Built Environ* 26:1018–1031
6. ISAZADE S, ALTAN MA (2023) Review of Literature on Measuring Energy Poverty. *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi* 24:336–361
7. Che X, Jiang M, Fan C (2021) Multidimensional Assessment and Alleviation of Global Energy Poverty Aligned With UN SDG 7. *Front Energy Res* 9
8. Faiella I, Lavecchia L (2021) Energy poverty. How can you fight it, if you can't measure it? *Energy Build* 233
9. Halkos GE, Gkampoura EC (2021) Coping with energy poverty: Measurements, drivers, impacts, and solutions. *Energies* (Basel). 14

10. Legendre B, Ricci O (2014) Measuring fuel poverty in France: Which households are the most fuel vulnerable? *Energy Econ* 49:620–628
11. Fizaine F, Kahouli S (2019) On the power of indicators: how the choice of fuel poverty indicator affects the identification of the target population. *Appl Econ* 51:1081–1110
12. Sareen S et al (2020) European energy poverty metrics: Scales, prospects and limits. *Glob Transit* 2
13. Li J, Li J, Guo K, Ji Q, Zhang D (2024) Policy spillovers from climate actions to energy poverty: international evidence. *Humanit Soc Sci Commun* 11:1106
14. Tundys B, Bretyn A, Inclusive Growth (2024) Energy Poverty and Digital and Social Development: Cross-Country Analysis of the European Union. *Energies (Basel)* 17:4830
15. Romero JC, Linares P, López X (2018) The policy implications of energy poverty indicators. *Energy Policy* 115:98–108
16. Sareen S et al (2020) European energy poverty metrics: Scales, prospects and limits. *Glob Transit* 2:26–36
17. Dubois U, Sinea A (2023) Methodological Challenges in Energy Poverty Research. *Int J Market Res* 65:340–358
18. Pereira G, González A, Ríos R (2021) Capturing Multidimensional Energy Poverty in South America: A Comparative Study of Argentina, Brazil, Uruguay, and Paraguay. *Front Sustainable Cities* 3
19. Thomson H, Snell C, Bouzarovski S (2017) Health, well-being and energy poverty in Europe: A comparative study of 32 European countries. *Int J Environ Res Public Health* 14
20. Wojewódzka-Wiewińska A, Dudek H, Ostasiewicz K (2024) Household Energy Poverty in European Union Countries: A Comparative Analysis Based on Objective and Subjective Indicators. *Energies (Basel)* 17:4889
21. Villalobos C, Chávez C, URIBE A (2020) Energy Poverty Measures and the Identification of the Energy Poor: A Comparison between the Utilitarian and Multidimensional Approaches in Chile. *SSRN Electron J*. <https://doi.org/10.2139/ssrn.3564827>
22. Bonatz N, Guo R, Wu W, Liu L (2019) A comparative study of the interlinkages between energy poverty and low carbon development in China and Germany by developing an energy poverty index. *Energy Build* 183:817–831
23. Papada L, Kaliampakos D (2017) Energy poverty in Greek mountainous areas: A comparative study. *J Mt Sci* 14
24. Aguilar JM, Ramos-Real FJ, Ramírez-Díaz AJ (2019) Improving indicators for comparing energy poverty in the Canary Islands and Spain. *Energies (Basel)*. 12
25. Dobrowolski Z, Adamišin P, Sługocki W, Kotylak S (2025) Energy Ladder, Decarbonisation and Energy Poverty: The European Union Inside. *Energies (Basel)* 18:1180
26. Zhu W et al (2025) Research on the process of energy poverty alleviation in China's provinces by new energy revolution from the perspective of time and space. *Energy* 322:135635

27. Thema J, Vondung F (2021) Expenditure-based indicators of energy poverty—an analysis of income and expenditure elasticities. *Energies* (Basel). 14
28. Fefferman N, Chen CF, Bonilla G, Nelson H, Kuo CP (2021) How limitations in energy access, poverty, and socioeconomic disparities compromise health interventions for outbreaks in urban settings. *iScience* 24
29. González-Eguino M (2015) Energy poverty: An overview. *Renew Sustain Energy Rev* 47:377–385
30. Nussbaumer P, Bazilian M, Modi V (2012) Measuring energy poverty: Focusing on what matters. *Renew Sustain Energy Rev* 16:231–243
31. Filho WL et al (2025) Addressing energy poverty: Regional trends and examples of best practice. *Energy Sustain Dev* 85:101647
32. Moore R (2012) Definitions of fuel poverty: Implications for policy. *Energy Policy* 49:19–26
33. Guevara Z, Espinosa-Aldama M, López-Corona O (2025) The evolution of energy poverty theory. *Energy Strategy Reviews* 61:101832
34. Kajoskoski T, Matschoss K, Heiskanen E, Laakso S (2025) Recent developments in the energy poverty and vulnerability research in Europe: a systematic literature review. *Energy Strategy Reviews* 61:101855
35. Martínez-Soto A et al (2025) Mapping energy poverty indices in urban scale: A case study of Santiago de Chile. *Energy Res Soc Sci* 127
36. Huang Z, Ming X, Duan H (2022) Heterogeneous effects of energy poverty convergence in Europe. *Energy Strategy Reviews* 41:100822
37. Liu J et al (2022) The role of Sustainable Development Goals to eradicate the multidimensional energy poverty and improve social Wellbeing's. *Energy Strategy Reviews* 42:100885
38. Tang J, Huang X, Chen HH, Gao W, Liu X (2024) Environmental regulation, energy poverty and social welfare: Based on China's provincial panel data. *Energy Strategy Reviews* 56:101592
39. He Y, Zhu Y, Zhu Y, Han D (2025) Research on the impact of clean energy demonstration provincial policy on energy poverty: Empirical evidence from China. *Energy Strategy Reviews* 59:101708
40. Flores Lanza M, Leonard A, Hirmer S (2024) Geospatial and socioeconomic prediction of value-driven clean cooking uptake. *Renew Sustain Energy Rev* 192:114199
41. Aguilar AJ, Guerrero-Rivera MF, de la Hoz-Torres ML (2025) Identifying urban energy-vulnerable areas: a machine learning approach. *J Building Eng* 109:113047
42. Casati P, Moner-Girona M, Khaleel SI, Szabo S, Nhamo G (2023) Clean energy access as an enabler for social development: A multidimensional analysis for Sub-Saharan Africa. *Energy Sustain Dev* 72:114–126
43. Ginocchi M, Ponci F, Monti A (2021) Sensitivity Analysis and Power Systems: Can We Bridge the Gap? A Review and a Guide to Getting Started. *Energies* (Basel) 14:8274

## Figures

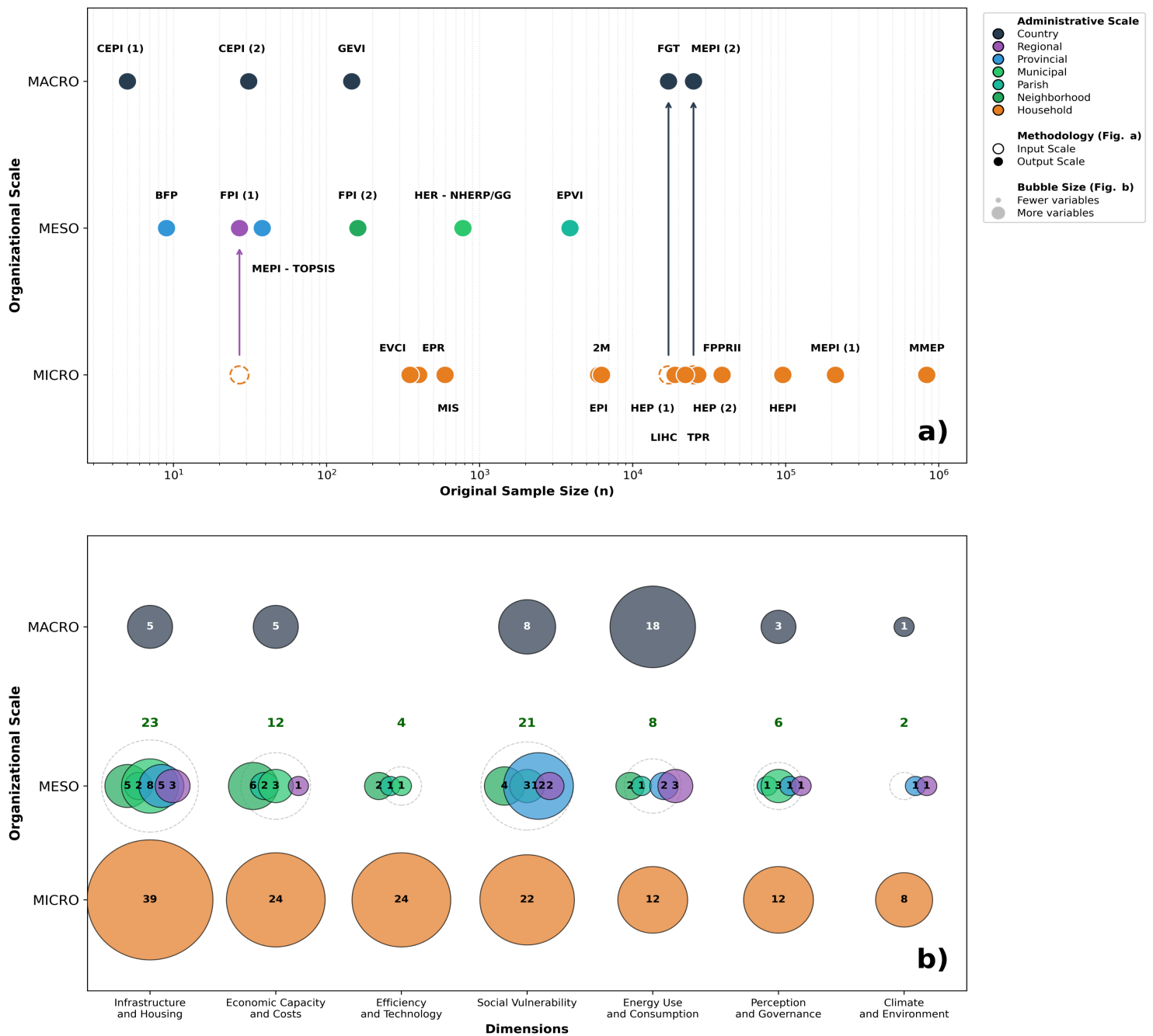


**Figure 1**

**Scalar asymmetry and dimensional contraction in energy poverty measurement. a)** Relationship between the original sample size (n) and the organizational scale of application. Arrows indicate indices with methodological transitions from micro-level inputs (households) to macro-level (national) outputs. **b)** Distribution of variables by dimension and administrative scale (Micro, Meso, Macro). The area of the bubbles represents the number of variables per subscale; dotted circles and green values indicate the aggregated total per dimension at the Meso level.

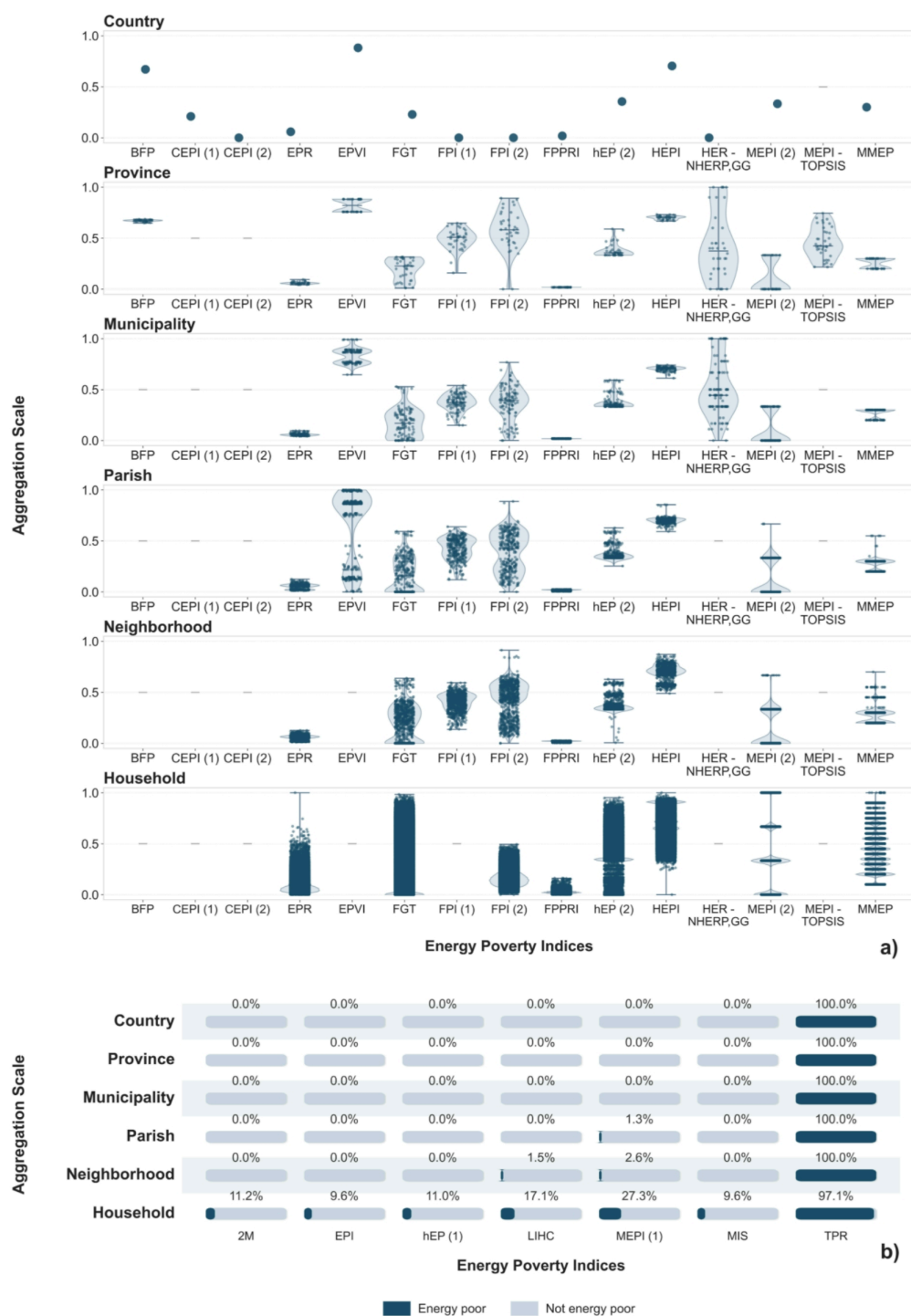


**Figure 2**

**Distribution of the 22 energy poverty indices across the six territorial aggregation scales (Household to Country). a)** Continuous indices ( $n = 15$ ), shown as violin plots with individual case values overlaid. **b)** Binary (dichotomous) indices ( $n = 7$ ), shown as the percentage of cases classified as energy poor (dark) versus not energy poor (light) at each scale.

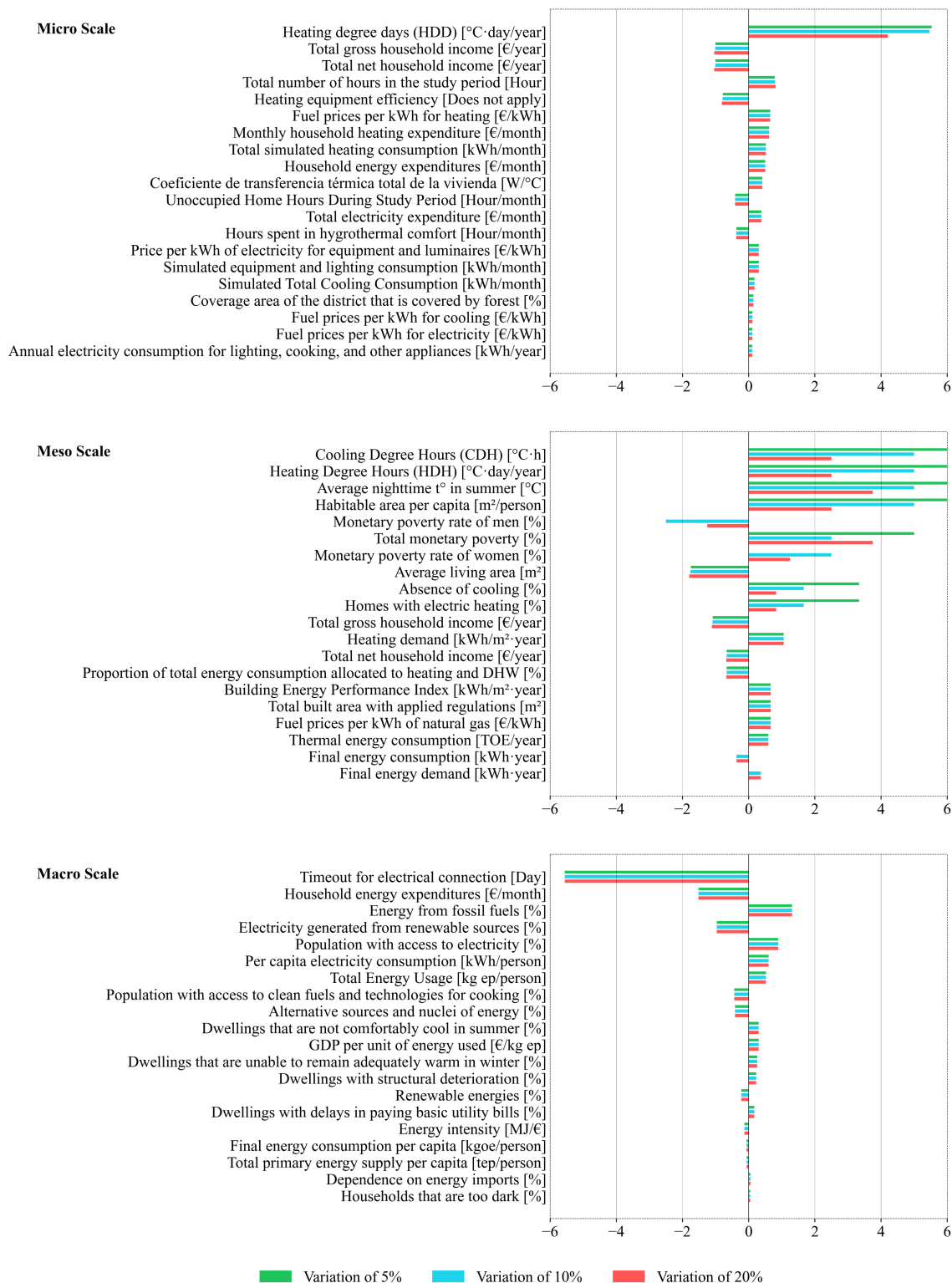


**Figure 3**

**Scalar dependence in the sensitivity hierarchy of energy poverty determinants.** Sensitivity analysis of continuous variables associated with continuous energy poverty indices across three aggregation scales. The Top-20 ranking was defined using the  $\pm 10\%$  perturbation scenario, while the  $\pm 5\%$  and  $\pm 20\%$  perturbations were incorporated to assess the robustness of the sensitivity hierarchy. The bars represent the magnitude and direction of the mean normalized sensitivity coefficient for each variable within each

scale. For readability, the x-axis was limited to the range -6 to 6; coefficients exceeding this range are discussed in the text.

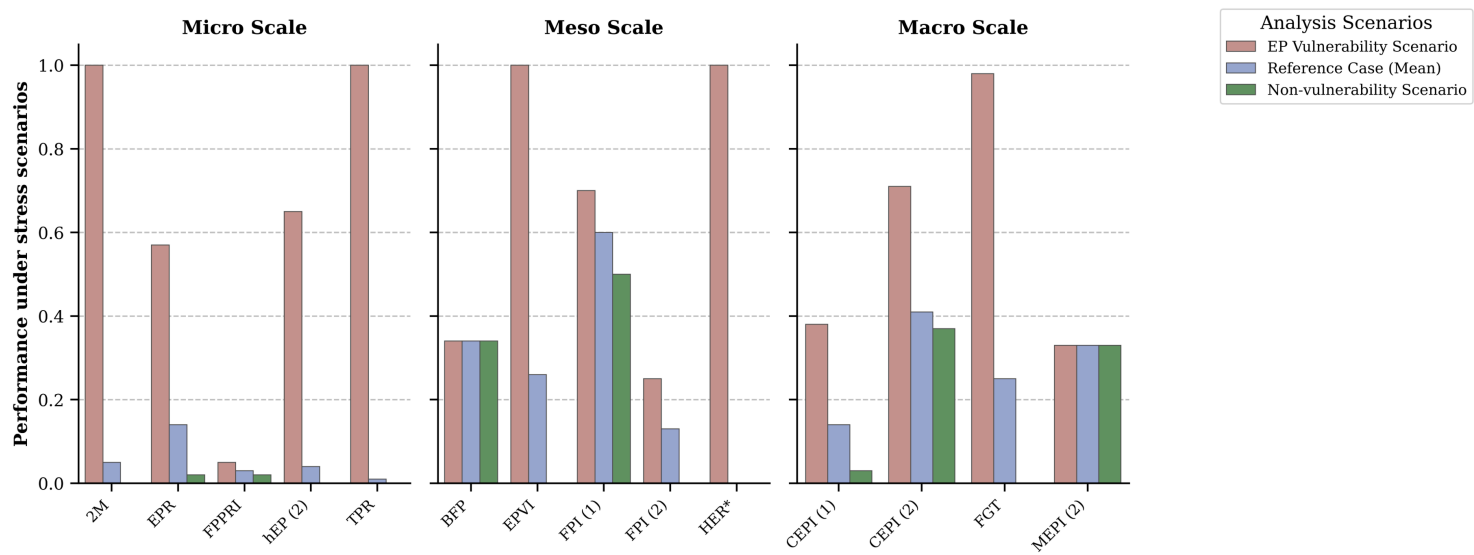


Figure 4

**Stress analysis of energy poverty indices.** Performance of indices at the micro, meso, and macro scales under three scenarios: vulnerability, reference case (mean), and non-vulnerability. The figure shows differences in sensitivity and discriminatory capacity across indices. While some indices cover the full range (0–1) and respond to extreme conditions, others remain within narrow or nearly invariant ranges, limiting their robustness and explanatory power.

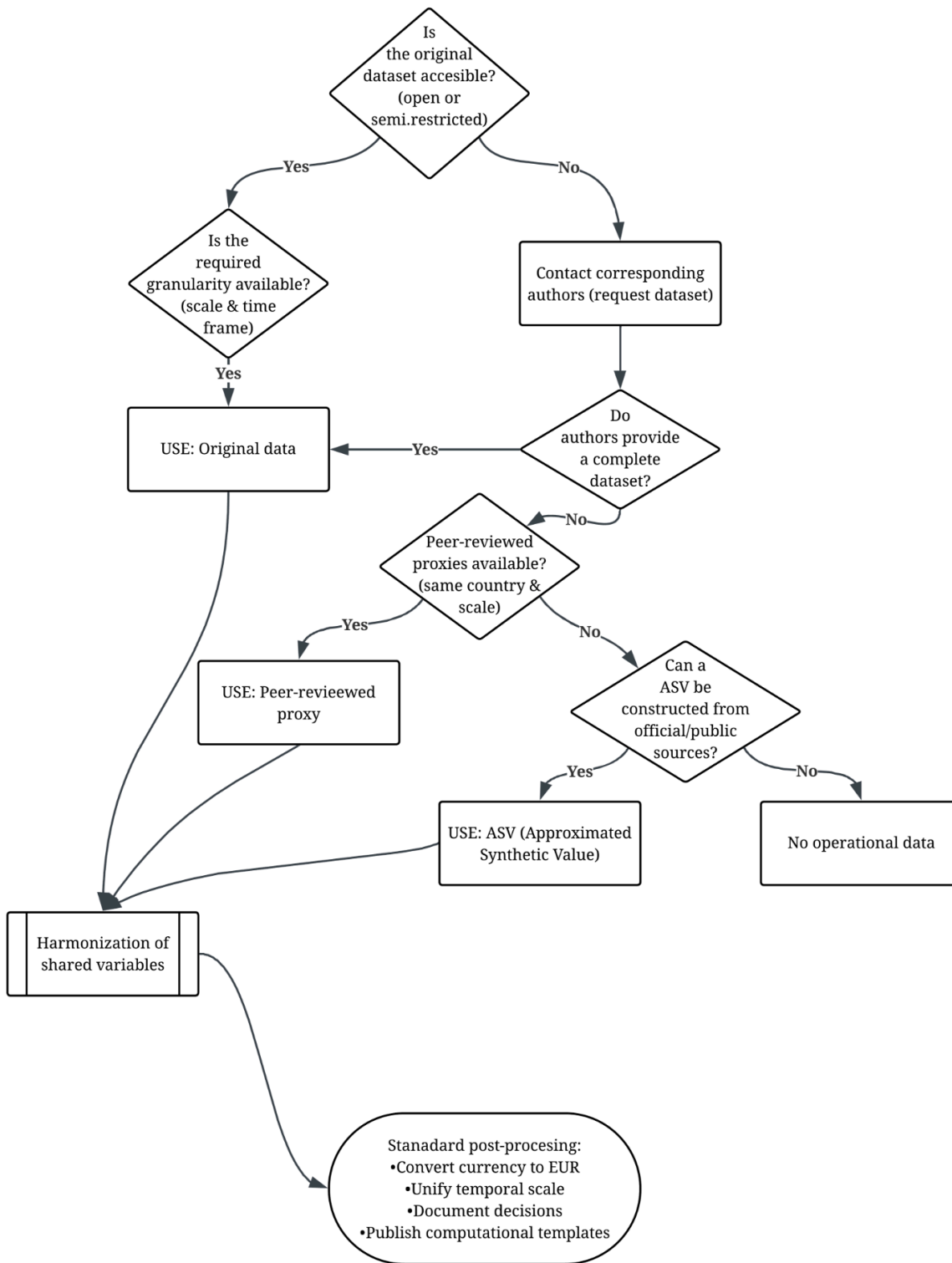


**Figure 5**

Hierarchical decision protocol for the retrieval, substitution, and harmonization of variables and input data used to reproduce energy poverty indices.